AD-A106 683 TEXAS UNIV AT AUSTIN CENTER FOR CYBERNETIC STUDIES F/G 12/1
QUASI-NEWTON METHODS CONVERGE AT THE GOLDEN SECTION RATE.(U)
AUG 81 J BARZILAI N00014-75-C-0569
UNCLASSIFIED CCS-RR-403 NL

1 OF 1
AD A
106683

END
DATE
FILMED
11-81
DTIC

CENTER FOR CYBERNETIC STUDIES

The University of Texas
Austin, Texas 78712

LEVEL II

(12)

Research Report CCS-403

QUASI-NEWTON METHODS CONVERGE AT THE GOLDEN SECTION RATE

by

J. Barzilai

August 1981

This research was partly supported by Project NR047-021, ONR Contract N00014-75-C-0569 with the Center for Cybernetic Studies, The University of Texas at Austin. Reproduction in whole or in part is permitted for any purpose of the United States Government.

CENTER FOR CYBERNETIC STUDIES

A. Charnes, Director
Business-Economics Building, 203E
The University of Texas at Austin
Austin, TX 78712
(512) 471-1821

Square root of 5

ABSTRACT

We prove that the rate of convergence of quasi-Newton methods is the golden section ratio $(1 + \sqrt{5})/2$.

KEY WORDS

Unconstrained minimization, Convergence rates, Quasi-Newton methods.

1. Introduction

Newton's method for the minimization of $f: R^n \to R$ requires computation and inversion of the Hessian matrix at each iteration. Quasi-Newton methods approximate the Hessian or its inverse by first order (i.e. gradient) information. These methods extend the classical secant (or False Position) method for $n > 1$ (see e.g. Luenberger [7]). They are known to converge to the solution superlinearly (see Dennis and Moré [3] and the references there). Thus, it is commonly accepted (e.g. [3]), that the price paid for the approximation of the Hessian by gradient information is a reduction from second order to superlinear convergence.

In [1,2], we developed new tools for the analysis of the rate of convergence of interpolatory algorithms. We use them in this paper to prove that actually, the rate of convergence of a class of quasi-Newton methods, without line-search and without restart, is given by the golden section ratio $(1+\sqrt{5})/2 \approx 1.618$. We note in passing that no other tools exist enabling one to establish convergence rates between superlinear and quadratic.

2. Rate of Convergence Analysis

Newton's method consists of the iteration $x_{k+1} = x_k - [\nabla^2 f(x_k)]^{-1} \cdot \nabla f(x_k)$. Here ∇f, $\nabla^2 f$ are the gradient and Hessian of f respectively and all vectors are column vectors. Quasi-Newton replace this equation with

(1) $$x_{k+1} = x_k - \alpha_k H_k \nabla f(x_k) ,$$

where the matrix H_k approximates the inverse of the Hessian, and the step-size $\alpha_k \in R$ is obtained by an exact or approximate line-search. The matrix H_k is required to satisfy

(2) $$H_{k+1}y_k = s_k ,$$

where

(3) $$y_k = \nabla f(x_{k+1}) - \nabla f(x_k) , \quad s_k = x_{k+1} - x_k .$$

For a thorough discussion of these methods see Dennis and Moré [3].

Henceforth, we will assume $\alpha_k = 1$ for all k, i.e., no line search is performed so that the iteration formula becomes

(4) $$x_{k+1} = x_k - H_k \nabla f(x_k) .$$

In the one dimensional case $(n = 1)$, equation (2) implies

$$H_k = \frac{x_k - x_{k-1}}{f'_k - f'_{k-1}}$$

with $f'_k = f'(x_k)$, so that (4) is the classical secant or False Position method (see Luenberger [7]). For this reason equation (2) is called the secant equation. Other names, e.g. quasi-Newton equation, are also in use. This equation plays a fundamental role in the classical theory of quasi-Newton methods as well as in our analysis.

The formulas expressing H_{k+1} in terms of H_k and the data are called updating formulas. Different updating formulas give rise to a variety of quasi-Newton methods. In addition, there are quasi-Newton methods which replace equations (2) and (4) with

(5) $$x_{k+1} = x_k - B_k^{-1} \nabla f(x_k) ,$$

(6) $$B_{k+1} s_k = y_k$$

together with an appropriate updating formula for the matrix B_k.

We recall our basic results on hyperosculatory interpolation algorithms developed in [1,2]. The interpolation algorithm studied there generates a sequence $\{x_k\}$ as follows. Let $s \geq 1$, $m \geq 0$ be fixed integers, and let $T: R^n \to R$ depend on $r = s(m+1)$ parameters. Given $m+1$ approximants $x_0, \ldots, x_{m+1}$ to the solution x^* of $\nabla f(x) = 0$, we use $x_{k-m}, \ldots, x_{k-1}, x_k$ to construct a new approximant x_{k+1}. First we interpolate f by T requiring

$$(7) \qquad T^{(i)}(x_{k-j}) = f^{(i)}(x_{k-j}) \quad j = 0, \ldots, m; \quad i = 0, \ldots, s-1 .$$

Here $f^{(1)} = \nabla f$, $f^{(2)} = \nabla^2 f$ etc. The new point x_{k+1} is determined by

$$(8) \qquad \nabla T(x_{k+1}) = 0 .$$

In [1], we proved that the sequence $\{x_k\}$, generated by this algorithm converges (locally) to the solution with Q- and R-rates of convergence at least p, where p is the unique positive solution of the equation $t^{m+1} - (s-1)t^m - s\sum_{j=0}^{m-1} t^j = 0$ (the sum is taken as zero if $m = 0$). For the definitions of the Q- and R-rates of convergence and their properties see [9, §9]. The derivation of this result is based on the analysis in Traub [11], where a difference relation for the errors $\|x_k - x^*\|$ is used to compute the rate.

To show that quasi-Newton methods as defined above can be regarded as interpolatory algorithms, we now characterize them by the requirements

$$(9) \qquad T(x_k) = f(x_k)$$

$$(10) \qquad \nabla T(x_k) = \nabla f(x_k)$$

$$(11) \qquad \nabla T(x_{k-1}) = \nabla f(x_{k-1}) ,$$

and

$$(12) \qquad \nabla T(x_{k+1}) = 0 ,$$

where T is the quadratic interpolation function

$$(13) \qquad T(x) = f(x_k) + (x-x_k)^T \nabla f(x_k) + \frac{1}{2}(x-x_k)^T B_k (x-x_k) ,$$

and where B_k is a symmetric nonsingular $n \times n$ matrix, and a^T stands for the transpose of the vector a.

Indeed, if T is defined by (13), equation (9) holds and

$$(14) \qquad \nabla T(x) = \nabla f(x_k) + B_k(x-x_k) ,$$

which implies (10). Using (14) in (12) we have $\nabla f(x_k) + B_k(x_{k+1}-x_k) = 0$, which is equivalent to (5). Finally the requirement (11) is equivalent to

$$\nabla f(x_k) + B_k(x_{k-1}-x_k) = \nabla f(x_{k-1}) ,$$

which is the secant equation (6).

So far we have interpreted all quasi-Newton algorithms as interpolatory algorithms. Note that (9)-(11) do not define hyperosculatory interpolation, since we do not require $T(x_{k-1}) = f(x_{k-1})$, therefore our results in [1] do not apply directly to the algorithm (9)-(12). For $n = 1$ the algorithm is precisely the secant method which is well known to have convergence order $(1 + \sqrt{5})/2$. We will now show that the rate of convergence of a class of quasi-Newton methods is induced by the underlying one-dimensional secant algorithm.

First we note that equation (9) is redundant. Indeed, equations (10)-(13) are sufficient to define the sequence $\{x_k\}$, for if $T(x)$ satisfies (9)-(13) and $T_1(x) = T(x) + a$ with $a \in R$, equation (9) may no longer hold for $T_1(x)$, but $\nabla T_1(x) = \nabla T(x)$ will produce the same value for x_{k+1}.

As in [1], we derive the basic difference equation we need by passing a curve in R^n through the points $x_{k-1}, x_k, x_{k+1}, x^*$, i.e., we determine a function $\psi: R \to R^n$ such that

$$(15)\qquad \begin{cases} \psi(t_{k-j}) = x_{k-j} & j=-1,0,1 \\ \psi(t^*) = x^* , & \end{cases}$$

where the parameter t is chosen so that

$$(16)\qquad t_{k-j} = \|x_{k-j}-x^*\| , \quad t^* = \|x^*-x^*\| = 0 .$$

This can evidently be done in infinitely many ways. We will later specify further restrictions on ψ. Defining $\bar{\theta}(t) = T(\psi(t))$, $\bar{\varphi}(t) = f(\psi(t))$ and $\theta(t) = \bar{\theta}'(t)$, $\varphi(t) = \bar{\varphi}'(t)$, we have from (10)-(12)

$$(17)\qquad \theta(t_k) = \varphi(t_k)$$

$$(18)\qquad \theta(t_{k-1}) = \varphi(t_{k-1})$$

$$(19)\qquad \theta(t_{k+1}) = 0$$

$$(20)\qquad \varphi(0) = 0 .$$

Having reduced the original equations to one-dimensional hyperosculatory interpolation ones, we are now able to derive a difference equation for the sequence $\{t_k\}$.

Theorem 1. If $\theta,\varphi \in C^{(2)}(J)$ where $J = \{t: |t| \le L\}$ for some $L > 0$, and if $t_{k-j} \in J$ $j=-1,0,1$ then equations (17)-(20) imply

$$(21)\qquad t_{k+1} = A_k\, t_k\, t_{k-1}$$

where

$$(22)\qquad A_k = \frac{\varphi^{(2)}(\xi)-\theta^{(2)}(\xi)}{2\theta'(\zeta)}$$

and ξ , ζ are in the interval spanned by t_{k-1}, t_k, t_{k+1} and 0.

Proof. By the remainder formula for a general interpolating function (see Ostrowski [10]), (17) and (18) imply

$$\varphi(t) - \theta(t) = \frac{\varphi^{(2)}(\xi(t)) - \theta^{(2)}(\xi(t))}{2}(t-t_k)(t-t_{k-1}) \tag{23}$$

with $\xi(t)$ in the interval spanned by t, t_k and t_{k-1}. By (19) we have $-\theta(0) = \theta(t_{k+1}) - \theta(0) = t_{k+1}\theta'(\zeta)$ with ζ between t_{k+1} and 0. Setting $t = 0$ in (23) and denoting $\xi = \xi(0)$ we therefore have

$$t_{k+1}\theta'(\zeta) = \frac{\varphi^{(2)}(\xi) - \theta^{(2)}(\xi)}{2}\, t_k\, t_{k-1}\,,$$

which completes the proof.

$\square$

Our main result now follows from equation (21).

<u>Theorem 2</u>. Let $f \in C^{(3)}$ in a neighborhood of the solution x^*. If $\nabla^2 f(x^*)$ is positive definite, and if the sequence $\{B_k\}$ is bounded, then there exists a neighborhood N of x^*, such that for all $x_0, x_1 \in N$, the sequence $\{x_k\}$ generated by the quasi-Newton algorithm converges to x^* with Q- and R-rates of convergence at least $(1+\sqrt{5})/2$.

<u>Proof</u>. This is an immediate consequence of the difference equation (21), if the sequence $\{A_k\}$ is bounded (see e.g. [6] or [11] and [2]).

Under the assumptions of the theorem and by definition of the functions θ, φ, it is therefore sufficient to show that the curve ψ can be chosen so that the derivatives of ψ are bounded at $t = 0$, and $\varphi'(0) \neq 0$.

Note that ψ is used to derive equation (21), but its construction is not a part of the algorithm. Assuming without loss of generality $\frac{\partial^2 f(x^*)}{\partial x_1^2} \neq 0$, and since $\varphi'(0) = \dot{\psi}(0)^T\ {}^2 f(x^*)\dot{\psi}(0)$, one can satisfy (15) and $\varphi'(0) \neq 0$ by choosing $\psi_i(t) = \sum_{j=0}^{r} a_{ji} t^j$ $(i = 1,\dots,n)$ with $a_{11} = 1$, $a_{1i} = 0$ $i = 2,\dots,n$. This completes the proof.

$\square$

Theorem 2 holds for all quasi-Newton methods. We now turn our attention to the so-called Broyden's class of quasi-Newton methods, which are defined by the updating formula

$$(24)\quad \begin{cases} H_{k+1} = H_k + \dfrac{s_k s_k^T}{s_k^T y_k} - \dfrac{H_k y_k y_k^T H_k}{y_k^T H_k y_k} + \alpha_k v_k v_k^T , \\ \text{with } y_k, s_k \text{ defined by (3),} \\ v_k = \left(y_k^T H_k y_k\right)^{\frac{1}{2}} \left[\dfrac{s_k}{s_k^T y_k} - \dfrac{H_k y_k}{y_k^T H_k y_k} \right] \end{cases}$$

and $\alpha_k \in [0,1]$.

Evidently boundedness of B_k and $H_k = B_k^{-1}$ is equivalent.

<u>Theorem 3</u>. Let $f \in C^{(3)}$ in a neighborhood of the solution x^*, and let $\nabla^2 f(x^*)$ be positive definite. If x_0, x_n are close enough to x^*, if H_0 is symmetric and positive definite, and if the matrices H_k are updated by (24), then $x_k \to x^*$ with Q- and R-rates of convergence at least $(1+\sqrt{5})/2$.

<u>Proof</u>. By the mean value theorem we have $y_k = A_k s_k$ where $A_k = \nabla^2 f(\bar{x})$ and $\bar{x}$ on the segment line connecting x_k and x_{k+1}. Fletcher [4] proved that the eigenvalues of $A_k^{\frac{1}{2}} H_k A_k^{\frac{1}{2}}$ are bounded. Since we assumed that $\nabla^2 f$ is continuous and positive definite at x^*, the eigenvalues of H_k are bounded and the result follows from Theorem 2.

$\square$

3. Concluding Remarks

Under traditional assumptions, we have proved that quasi-Newton methods inherit their rate of convergence from the underlying secant method (cf. Luenberger [6, §7.2].

Thus, the assumption in Theorem 8.9 of [3] that equation (8.21) of that paper holds, is not made here. Similarly, no assumption has been made on the linear independence of the directions $\{s_k\}$ (cf. Moré and Trangenstein [8]).

We have not broadened our analysis to quasi-Newton methods beyond those belonging to Broyden's class of updates (and their inverse updated in the sense of [3]), in order not to obscure the main points in our analysis. The well known Davidon-Fletcher-Powell and Broyden-Fletcher Goldfarb-Shanno algorithms fall in this category. While the latter algorithm is the best available at present, our analysis in [1] suggests that faster algorithms can be designed utilizing gradient information only.

Our results extend with the obvious modifications for the problem of solving $F(x) = 0$, $F: R^n \to R^n$ discussed in the first part of [3]. They also extend to the infinite dimensional case if the coefficients A_k in the basic difference equation (21) are bounded.

From our point of view, the rate of convergence of quasi-Newton methods has nothing to do with their so-called quadratic termination property. It is a consequence of the <u>data</u> used in the interpolatory equations (7) (see [1,2]). Therefore, the Huang class of updates [5] is too wide in the sense that it contains updates which do not satisfy the secant equation. Note also that Theorem 8.10 of [3] is not interesting in the sense that $1.6^n > 2$ for all $n > 1$.

Finally, note that the common observation that Newton's method is self corrective in the sense that x_{k+1} depends explicitly on x_k only, while quasi-Newton methods carry along bad effects from previous iterations, is not justified. The fact that quasi-Newton methods are two-point interpolatory algorithms, is exactly their advantage over Newton's method (see [10, §6.4], [1] and [2]).

REFERENCES

[1] J. Barzilai, Unconstrained minimization by interpolation: Rates of Convergence, Research Report 389, Center for Cybernetic Studies, The University of Texas at Austin, 1980.

[2] J. Barzilai and A. Ben-Tal, Nonpolynomial and inverse interpolation for line search: synthesis and convergence rates, Research Report 385, Center for Cybernetic Studies, The University of Texas at Austin, 1980.

[3] J.E. Dennis and J.J. Moré, Quasi-Newton methods, motivation and theory, SIAM Review, 19(1977), pp. 46-89.

[4] R. Fletcher, A new approach to variable metric algorithms, Comput. J., 13(1970), pp. 317-322.

[5] H.Y. Huang, Unified approach to quadratically convergent algorithms for function minimization, J. Optimization Theory Appl., 5(1970), pp. 405-423.

[6] E. Isaacson and H.B. Keller, Analysis of Numerical Methods, Wiley, New York, 1966.

[7] D.G. Luenberger, Introduction to Linear and Nonlinear Programming, Addison-Wesley, Reading, Mass., 1973.

[8] J.J. Moré and J.A. Trangenstein, On the global convergence of Broyden's method, Math. Comp., 30(1976), pp. 523-540.

[9] J.M. Ortega and W.C. Rheinboldt, Iterative Solution of Nonlinear Equations in Several Variables, Academic Press, New York, 1970.

[10] A.M. Ostrowski, Solution of Equations and Systems of Equations, 2nd ed., Academic Press, New York, 1966.

[11] J.F. Traub, Iterative Methods for the Solution of Equations, Prentice-Hall, Englewood Cliffs, N.J., 1964.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER CCS 403	2. GOVT ACCESSION NO. AD-A106683	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Quasi-Newton Methods Converge at the Golden Section Rate		5. TYPE OF REPORT & PERIOD COVERED 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) J. Barzilai		8. CONTRACT OR GRANT NUMBER(s) N00014-75-C-0569
9. PERFORMING ORGANIZATION NAME AND ADDRESS Center for Cybernetic Studies, UT Austin Austin, TX 78712		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research (Code 434) Washington, DC		12. REPORT DATE August 1981 13. NUMBER OF PAGES 11
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

This document has been approved for public release and sale; its distribution is unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Unconstrained minimization, Convergence rates, Quasi-Newton methods

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

We prove that the rate of convergence of quasi-Newton methods is the golden section ratio $(1+\sqrt{5})/2$.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

AT

LM